

REMARKS

In response to final Office Action dated October 9, 2008, claims 1, 14, 20, and 24 have been amended and claims 2-4, 17, 18, 21, 28, and 30-60 have been canceled. Therefore, claims 1, 5-16, 19, 20, and 22-29 remain in the case. The Applicants respectfully request that this amendment be entered under 37 C.F.R. 1.116 to place the above-referenced application in condition for allowance or, alternatively, in better condition for appeal. In light of the amendments and arguments set forth herein, reexamination and reconsideration of the application are requested.

Section 103(a) Rejections

The final Office Action rejected claims 1, 6-10, 12-14, 16, 20, and 22-27 under 35 U.S.C. § 103(a) as being unpatentable over a paper by Sturim et al. entitled "Speaker Indexing in Large Audio Databases Using Anchor Models" in view of a paper by Waibel et al. entitled "Phoneme Recognition Using Time-Delay Neural Networks", and further in view of Hermansky et al (U.S. Patent No. 7,254,538). The Office Action contended that the combination of Sturim et al., Waibel et al. and Hermansky et al. teaches all the elements of the Applicants' claimed invention.

In response, the Applicants respectfully traverse these rejections. In general, the Applicants submit that the combination of Sturim et al., Waibel et al. and Hermansky et al. is lacking several elements of the Applicants' claimed invention. More specifically, neither Sturim et al., Waibel et al., nor Hermansky et al. disclose, either explicitly or implicitly, the material claimed features of:

1. (Recited in amended independent claim 1): "obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models;"

2. (Recited in amended independent claim 14): "obtaining a preliminary output of the plurality of anchor models from the convolutional neural network during training of the discriminatively-trained classifiers before final nonlinearities are applied by the second layer in order to generate a modified feature vector output;"
3. (Recited in amended independent claim 20): "obtaining a preliminary output of the plurality of anchor models from a time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models;"
4. (Recited in amended independent claim 24): "obtaining during training the plurality of anchor model outputs from the convolutional neural network prior to application of final nonlinearities by the second layer to generate a modified plurality of anchor model outputs;"

Further, the combination of Sturim et al., Waibel et al., and Hermansky et al. fails to appreciate the advantages of these claimed features. In addition, there is no technical suggestion or motivation disclosed in either Sturim et al., Waibel et al., or Hermansky et al. to define these claimed features. Thus, the Applicants submit that the combination of Sturim et al., Waibel et al., and Hermansky et al. cannot make obvious the Applicants' claimed features listed above.

To make a *prima facie* showing of obviousness, all of the claimed features of an Applicant's invention must be considered, especially when they are missing from the prior art. If a claimed feature is not disclosed in the prior art and has advantages not appreciated by the prior art, then no *prima facie* showing of obviousness has been made. The Federal Circuit Court has held that it was an error not to distinguish claims over a combination of prior art references where a material limitation in the claimed system and its purpose was not taught therein. *In re Fine*, 837 F.2d 1071, 5 USPQ2d

1596 (Fed. Cir. 1988). Moreover, as stated in the MPEP, if a prior art reference does not disclose, suggest or provide any motivation for at least one claimed feature of an Applicant's invention, then a *prima facie* case of obviousness has not been established (MPEP § 2142).

Amended Independent Claims 1, 14, 20, and 24

Amended independent claim 1 recites a method for processing audio data. The method includes training time-delay neural network (TDNN) classifiers using a time-delay neural network that uses a first layer followed by a second layer having a nonlinearity, using discriminatively-trained classifiers that are time-delay neural network classifiers to produce a plurality of anchor models, and applying the plurality of anchor models to the audio data. The method also includes obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models, normalizing the output of the plurality of anchor models to generate a normalized output of the plurality of anchor models, mapping the normalized output of the plurality of anchor models into frame tags, and producing the frame tags.

Amended independent claim 14 recites a computer-implemented process for processing audio data. The method includes applying a plurality of anchor models to the audio data, the plurality of anchor models comprising discriminatively-trained classifiers of a convolutional neural network that were previously trained using a training technique using a first layer followed by a second layer having a nonlinearity, and obtaining a preliminary output of the plurality of anchor models from the convolutional neural network during training of the discriminatively-trained classifiers before final nonlinearities are applied by the second layer in order to generate a modified feature vector output. The method also includes normalizing the modified feature vector output to generate normalized anchor model output, mapping the normalized anchor model output into frame tags, and producing the frame tags.

Amended independent claim 20 recites a method for processing audio data containing a plurality of speakers. The method includes training time-delay neural network (TDNN) classifiers using a time-delay neural network that uses a first layer followed by a second layer having a nonlinearity, using the TDNN classifiers to produce a plurality of anchor model outputs, and applying the plurality of anchor models to the audio data. The method also includes obtaining a preliminary output of the plurality of anchor models from a time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models, normalizing the output of the plurality of anchor models to generate a normalized output of the plurality of anchor models, and mapping the normalized output of the plurality of anchor models into frame tags. The method further includes constructing a list of start and stop times for each of the plurality of speakers based on the frame tags, where the discriminatively-trained classifiers were previously trained using a training set containing a set of training speakers, and where the plurality of speakers is not in the set of training speakers.

Amended independent claim 24 recites a computer-readable medium having computer-executable instructions for processing audio data. The instructions include training discriminatively-trained classifiers that are time-delay neural network (TDNN) classifiers in a discriminative manner on a convolutional neural network using a training technique such that the training occurs during a training phase to generate parameters that can be used at a later time by the TDNN classifiers and includes two layers with a first layer including a one-dimensional convolution followed by a second layer having a nonlinearity, and using the TDNN classifiers to produce a plurality of anchor model outputs. The instructions also include obtaining during training the plurality of anchor model outputs from the convolutional neural network prior to application of final nonlinearities by the second layer to generate a modified plurality of anchor model outputs, normalizing the modified plurality of anchor model output to generate normalized anchor model outputs, and clustering the normalized anchor model outputs into frame tags of speakers that are contained in the audio data.

Amended claims 1, 14, 20, and 24 each contain the feature that a preliminary output is obtained from a neural network during training of classifiers and before any nonlinearities are applied. The Applicants' specification states that the "normalization module 400 initially accepts the convolutional neural network outputs 600. These outputs 600 are obtained prior to an application of the final nonlinearity process. In other words, during training, the convolutional neural network uses nonlinearities, but the normalization module 400 obtains the output 600 before the final nonlinearities are applied" (specification, page 19, lines 13-17). Moreover, "the normalization process begins by accepting anchor model outputs before the final non-linearity of the convolutional neural network" (specification, page 24, lines 7-10).

For example, in the working example presented in the specification, the "TDNN classifier 1415 has two layers with each layer including a one-dimensional convolution followed by a nonlinearity" (specification, page 28, lines 29-30). This includes "omitting the nonlinearity contained in the second layer of the TDNN classifier 1415 (in this case the TDNN classifier was trained using the cross-entropy technique). In other words, the numbers before the nonlinearity are used (there were 76 of these numbers)" (specification, page 30, lines 25-29). Thus, normalization is performed using output that is obtained before the second layer having the nonlinearity is applied to that output.

In contrast, neither Sturim et al. nor Waibel et al. disclose "obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models.

The Office Action stated that Hermansky et al. disclose "a system where the final nonlinearity is omitted at column 2, lines 23-25 and at column 3, lines 32-35." However, Hermansky et al. merely teach omitting the final nonlinearity in the output layer of the neural network after training has occurred. In particular, Hermansky et al. state that "the present invention transforms the output of one or more neural networks that are trained to derive subword (phone) . . ." (col. 2, lines 16-18). Hermansky et al. then go on to say

that "such warping includes omitting the output layer of the neural network trained using softmax nonlinearity" (col. 2, lines 23-25). In other words, the omission of the nonlinearity occurs after training, not during training.

Moreover, Hermansky et al. state that ". . . original features 10 derived from an audio stream are input to a neural network such as a multi-layer perceptron (MLP) 12 phone classifier trained to estimate subword (phone) . . ." (col. 3, lines 15-18). Hermansky et al. then go on to say that "[A]lternatively, the final nonlinearity in the output layer of the neural network MLP 12 may be omitted" (col. 3, lines 32-33). Again, the omission occurs after training, not during training.

In addition, the combination of Sturim et al., Waibel et al., and Hermansky et al. also fails to appreciate or recognize the advantages of this feature. In particular, this feature is part of a normalization process, which "is used to remove spurious discrepancies caused by scaling by mapping data to a unit sphere" (specification, page 24, lines 7-8). Neither Sturim et al., Waibel et al., nor Hermansky et al. discuss or appreciate these advantages of this feature recited in the Applicants' amended claims 1, 14, 20, and 24.

The Applicants, therefore, submit that obviousness cannot be established since the combination of Sturim et al., Waibel et al. and Hermansky et al. fails to teach, disclose, suggest or provide any motivation for the features recited in amended claims 1, 14, 20, and 24, as discussed above. In addition to explicitly lacking these features, the combination of Sturim et al., Waibel et al., and Hermansky et al. fails to implicitly disclose, suggest, or provide motivation for these features. Further, the combination also fails to appreciate advantages of these claimed features.

Therefore, as set forth in *In re Fine* and MPEP § 2142, the combination of Sturim et al., Waibel et al., and Hermansky et al. cannot render amended independent claims 1, 14, 20, and 24 obvious because Sturim et al., Waibel et al., and Hermansky et al. are missing the material features recited in claims 1, 14, 20, and 24, as discussed above.

Consequently, because a *prima facie* case of obviousness cannot be established due to the lack of "some teaching, suggestion, or incentive supporting the combination", the rejections must be withdrawn. ACS Hospital Systems, Inc. v. Montefiore Hospital, 732 F.2d 1572, 1577, 221 USPQ 929, 933 (Fed. Cir. 1984); MPEP 2143.01.

Accordingly, the Applicants respectfully submit that amended independent claims 1, 14, 20, and 24 are patentable under 35 U.S.C. § 103(a) over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. based on the amendments to claims 1, 14, 20, and 24, and the legal and technical arguments set forth above and below. Moreover, claims 6-10, 12, and 13 depend from amended independent claim 1, claim 16 depends from amended independent claim 14, claims 22 and 23 depend from amended independent claim 20, and claims 25-27 depend from amended independent claim 24, and are also nonobvious over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. (MPEP § 2143.03). The Applicants, therefore, respectfully request reexamination, reconsideration and withdrawal of the rejection of claims 1, 6-10, 12-14, 16, 20, and 22-27 under 35 U.S.C. § 103(a) as being unpatentable over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al.

The Office Action rejected claims 5 and 15 under 35 U.S.C. § 103(a) as being unpatentable over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. as applied to claims 1 and 14 above, and further in view of a paper by Lavagetto entitled "Time-Delay Neural Network for Estimating Lip Movements form Speech Analysis". The Office Action contended that the combination of Sturim et al., Waibel et al., Hermansky et al., and Lavagetto teach all the elements recited in these claims.

In response, the Applicants respectfully traverse these rejections. In particular, the Applicants submit that the combination of Sturim et al., Waibel et al., Hermansky et al., and Lavagetto is lacking several elements of the Applicants' claimed invention. More specifically, neither Sturim et al., Waibel et al., Hermansky et al., nor Lavagetto disclose, either explicitly or implicitly, the material claimed features of:

1. (Recited in amended independent claim 1): "obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models;"
2. (Recited in amended independent claim 14): "obtaining a preliminary output of the plurality of anchor models from the convolutional neural network during training of the discriminatively-trained classifiers before final nonlinearities are applied by the second layer in order to generate a modified feature vector output."

Further, the combination fails to appreciate the advantages of these claimed features. In addition, there is no technical suggestion or motivation disclosed in either Sturim et al., Waibel et al., Hermansky et al., or Lavagetto to define these claimed features. Thus, the Applicants submit that the combination of Sturim et al., Waibel et al., Hermansky et al., and Lavagetto cannot make obvious the Applicants' claimed features listed above.

Regarding the features recited in claims 1 and 14, it was argued above that neither Sturim et al., Waibel et al., nor Hermansky et al., alone or in combination, disclose these features.

Lavagetto adds nothing to the cited combination that would render obvious Applicants' amended claims 1 and 14. In particular, Lavagetto merely discloses using a time-delay neural network to analyze speech to estimate lip movements. Nowhere, however, does Lavagetto teach the Applicant's claimed feature recited in claim 1 of "obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models" or the

feature recited in claim 14 of "obtaining a preliminary output of the plurality of anchor models from the convolutional neural network during training of the discriminatively-trained classifiers before final nonlinearities are applied by the second layer in order to generate a modified feature vector output." In addition, Lavagetto fails to appreciate or recognize the advantages of these claimed features.

The Applicants, therefore, submit that obviousness cannot be established since the combination of Sturim et al., Waibel et al., Hermansky et al., and Lavagetto fails to teach, disclose, suggest or provide any motivation for the Applicants' claimed features recited in claims 1 and 14. In addition to explicitly lacking these features, Sturim et al., Waibel et al., Hermansky et al., and Lavagetto fail to implicitly disclose, suggest, or provide motivation for these features. Further, the combination also fails to appreciate the advantages of these claimed features.

Therefore, as set forth in *In re Fine* and MPEP § 2142, the combination of Sturim et al., Waibel et al., Hermansky et al., and Lavagetto cannot render the Applicants' claims 1 and 14 obvious. Consequently, because a *prima facie* case of obviousness cannot be established due to the lack of "some teaching, suggestion, or incentive supporting the combination", the rejection must be withdrawn. ACS Hospital Systems, Inc. v. Montefiore Hospital, 732 F.2d 1572, 1577, 221 USPQ 929, 933 (Fed. Cir. 1984); MPEP 2143.01.

Accordingly, the Applicants respectfully submit that amended independent claims 1 and 14 are patentable under 35 U.S.C. § 103(a) over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. as applied to claims 1 and 14, and in view of Lavagetto based on the amendments to claims 1 and 14 and the legal and technical arguments set forth above and below. Moreover, claim 5 depends from amended independent claim 1, and claim 15 depends from amended independent claim 14, and are also nonobvious over the cited art (MPEP § 2143.03). The Applicants, therefore, respectfully request reexamination, reconsideration and withdrawal of the rejection of claims 5 and 15.

The Office Action rejected claims 11, and 29 under 35 U.S.C. § 103(a) as being unpatentable over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. as applied to claims 10 and 25 above, and further in view of Liu (U.S. Patent No. 6,615,170). The Office Action contended that the combination of Sturim et al., Waibel et al., Hermansky et al., and Liu teach all the elements recited in these claims.

In response, the Applicants respectfully traverse these rejections. Specifically, the Applicants submit that the combination of Sturim et al., Waibel et al., Hermansky et al., and Liu is lacking several elements of the Applicants' claimed invention. More specifically, neither Sturim et al., Waibel et al., Hermansky et al., nor Liu disclose, either explicitly or implicitly, the material claimed features of:

1. (Recited in amended independent claim 1): "obtaining a preliminary output of the plurality of anchor models from the time-delay neural network during training of the TDNN classifiers before final nonlinearities are applied by the second layer in order to generate an output of the plurality of anchor models;"
2. (Recited in amended independent claim 24): "obtaining during training the plurality of anchor model outputs from the convolutional neural network prior to application of final nonlinearities by the second layer to generate a modified plurality of anchor model outputs;"

Further, the combination fails to appreciate the advantages of these claimed features. In addition, there is no technical suggestion or motivation disclosed in either Sturim et al., Waibel et al., Hermansky et al., or Liu to define these claimed features. Thus, the Applicants submit that the combination of Sturim et al., Waibel et al., Hermansky et al., and Liu cannot make obvious the Applicants' claimed features listed above.

Regarding the features recited in claims 1 and 24, it was argued above that neither Sturim et al., Waibel et al., nor Hermansky et al., alone or in combination, disclose these features.

Liu adds nothing to the cited combination that would render obvious Applicants' claims 1 and 24. Nowhere does Liu teach the Applicant's claimed feature recited in claim 1 and recited in claim 24. In addition, Liu fails to appreciate or recognize the advantages of these claimed features.

The Applicants, therefore, submit that obviousness cannot be established since the combination of Sturim et al., Waibel et al., Hermansky et al., and Liu fails to teach, disclose, suggest or provide any motivation for the Applicants' claimed features recited in claims 1 and 24. In addition to explicitly lacking these features, Sturim et al., Waibel et al., Hermansky et al., and Liu fail to implicitly disclose, suggest, or provide motivation for these features. Further, the combination also fails to appreciate the advantages of these claimed features.

Therefore, as set forth in *In re Fine* and MPEP § 2142, the combination of Sturim et al., Waibel et al., Hermansky et al., and Liu cannot render the Applicants' claims 1 and 24 obvious. Consequently, because a *prima facie* case of obviousness cannot be established due to the lack of "some teaching, suggestion, or incentive supporting the combination", the rejection must be withdrawn. ACS Hospital Systems, Inc. v. Montefiore Hospital, 732 F.2d 1572, 1577, 221 USPQ 929, 933 (Fed. Cir. 1984); MPEP 2143.01.

Accordingly, the Applicants respectfully submit that amended independent claims 1 and 24 are patentable under 35 U.S.C. § 103(a) over Sturim et al. in view of Waibel et al. and further in view of Hermansky et al. as applied to claims 10 and 25, and in view of Liu based on the amendments to claims 1 and 24 and the legal and technical arguments set forth above and below. Moreover, claim 11 depends from amended independent claim 1, and claim 29 depends from amended independent claim 24, and are also nonobvious over the cited art (MPEP § 2143.03). The Applicants, therefore, respectfully request

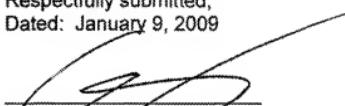
reexamination, reconsideration and withdrawal of the rejection of claims 11 and 29.

Conclusion

In view of the amendments to claims 1, 14, 20, and 24, and the arguments set forth above, the Applicants submit that pending claims 1, 5-16, 19, 20, and 22-29 are in condition for immediate allowance. The Examiner, therefore, is respectfully requested to withdraw the outstanding rejections of the claims and to pass all of the pending claims of this application to issue.

In an effort to expedite and further the prosecution of the subject application, the Applicants kindly invite the Examiner to telephone the Applicants' attorney at (805) 278-8855 if the Examiner has any comments, questions or concerns, wishes to discuss any aspect of the prosecution of this application, or desires any degree of clarification of this response.

Respectfully submitted,
Dated: January 9, 2009



Craig S. Fischer
Registration No. 42,535
Attorney for Applicants

LYON & HARR, L.L.P.
300 East Esplanade Drive, Suite 800
Oxnard, CA 93036-1274
Tel: (805) 278-8855
Fax: (805) 278-8064